CLAIMS

1. A method of logging updates in a main-memory transaction-processing system having main memory for storing a database, one or more log disks for storing log records for parallel recovery of the main memory database, and one or more backup disks for storing a copy of the main memory database, the method comprising the steps of:

taking a before image of the database before an update to the database is made;

taking an after image of the database after the update is made;

generating a differential log as a log body of each log record by applying a bit-wise exclusive-OR (XOR) operation between the before image and the after image;

recovering from a failure by applying the XOR operation between the differential log and the before-image.

2. The method of Claim 1, wherein the database comprises a plurality of fixed-size pages.

3. The method of Claim 2, wherein said each log record has a log header comprising:

LSN (Log Sequence Number) for storing log sequence;

TID(Transaction ID) for storing the identity of the transaction that created the log record;

Previous LSN for storing the identity of the most recently created log by the same transaction;

Type for storing the type of the log record;

Backup ID for storing the relation between the log record and the updated page for use with fuzzy checkpointing;

Page ID for storing the identity of an updated page;

Offset for storing the starting offset of an updated area within the updated page; and

Size for storing the size of the updated area.

4. The method of Claim 1, further comprising the step of:

checkpointing by occasionally writing the database in the main memory to said one or more backup disks as backup data.

5. The method of Claim 4, wherein the step of checkpointing uses the transaction

consistent checkpointing policy.

6. The method of Claim 4, wherein the step of checkpointing uses the action consistent checkpointing policy.

7. The method of Claim 4, wherein the step of checkpointing uses the fuzzy checkpointing policy.

8. The method of Claim 4, wherein the step of recovering comprises the steps of:

loading the backup data from said one or more backup disks into the main memory database; and

loading the log from said one or more log disks into the main memory database in order to restore the main memory database to the most recent consistent state.

9. The method of Claim 8, wherein the step of loading the backup data is executed in parallel by partitioning the backup data.

10. The method of Claim 8, wherein said step of loading the log comprises the steps of:

reading the log records from said one or more log disks; and

playing the log records in two pass to restore the main memory database to the latest consistent state.

11. The method of Claim 10, wherein the step of reading the log records and the step of playing the log records are executed in a pipeline.

12. The method of Claim 10, wherein the step of reading the log records is executed in parallel by partitioning the log records as well as the step of playing the log records.

13. The method of Claim 12, wherein the step of reading the log records and the step of playing the log records are executed in a pipeline.

14. The method of Claim 8, wherein the step of loading the log comprises the steps of:

reading log records from said one or more log disks; and

playing the log records in one pass to restore the main memory database to the latest consistent state.

15. The method of Claim 14, wherein the step of reading the log records and the step of playing the log records are executed in a pipeline.

16. The method of Claim 14, wherein the step of reading the log records is executed in parallel by partitioning the log records as well as the step of playing the log records.

17. The method of Claim 16, wherein the step of reading the log records and the step of playing the log records are executed in a pipeline.

18. The method of Claim 8, further comprising the step of filling the main memory database with 0s in advance.

19. The method of Claim 18, wherein the step of loading the backup data comprises the steps of:

reading the backup data from said one or more backup disks; and

playing the backup data by applying the XOR operation between the backup data and the main memory database.

20. The method of Claim 19, wherein the step of reading the backup data and the step of playing the backup data are executed in a pipeline.

21. The method of Claim 19, wherein the step of reading the backup data is executed in parallel by partitioning the backup data as well as the step of playing the backup data.

22. The method of Claim 21, wherein the step of reading the backup data and the step of playing the backup data are executed in a pipeline.

23. The method of Claim 19, wherein the step of loading the backup data and the step of loading the log records are executed in parallel.

24. A transaction processing system allowing recovery from a failure, comprising

main memory for storing a database,

one or more log disks for storing log records for parallel recovery of the main memory database,

one or more backup disks for storing a copy of the main memory database;

means for generating a differential log as part of the log body by applying a bit-wise exclusive-OR (XOR) between a before-image of the database before an update to the database is made and an after-image of the database after the update is made; and

means for recovering from a failure by applying the XOR operation between the differential log and the before-image.

25. The system of Claim 24, wherein the database comprises a plurality of fixed-size pages.

26. The system of Claim 24, further comprising:

means for checkpointing by occasionally writing the database in the main memory to one or more backup disks as backup data; and

27. The system of Claim 26, wherein the means for checkpointing uses the transaction consistent checkpointing policy.

28. The system of Claim 26, wherein the means for checkpointing uses the action consistent checkpointing policy.

29. The system of Claim 26, wherein the means for checkpointing uses the fuzzy checkpointing policy.

30. The system of Claim 26, wherein the means for recovering comprises:

means for loading the backup data into the main memory database; and

means for loading the log into the main memory database.

[system + checkpointing + BL/LL + BR/BP]

31. The system of Claim 30, wherein the means for loading the backup data comprises:

means for reading the backup data from one or more backup disks; and

means for playing the backup data to restore the main memory database to the state when the backup was made by applying the XOR operation between the backup data and the main memory database.

32. The system of Claim 30, wherein the means for loading the log comprises:

means for reading the log records from the log disk; and

means for playing the log records in two pass to restore the main memory database to the latest consistent state.

33. The system of Claim 30, wherein the means for loading the log comprises:

means for reading the log records from the log disk; and

means for playing the log records in one pass to restore the main memory database to the latest consistent state.

34. A computer-readable storage medium that contains a program for logging updates in a main-memory transaction-processing system having main memory for storing a database, one or more log disks for storing log records for parallel recovery of the main memory database, and one or more backup disks for storing a copy of the main memory database, where the program under the control of a CPU performs the steps of:

taking a before image of the database before an update to the database is made;

taking an after image of the database after the update is made;

generating a differential log as a log body of each log record by applying a bit-wise exclusive-OR (XOR) operation between the before image and the after image;

recovering from a failure by applying the XOR operation between the differential log and the before-image.

35. The storage medium of Claim 34, wherein the medium is a CD.

37. The storage medium of Claim 34, wherein the medium is a magnetic tape.